LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-406927

Dynamically Tunable Memory in Two-Component Gene Circuit

C.-M. Ghim, E. Almaas

September 11, 2008

Physical Review Letters

Disclaimer

Dynamically Tunable Memory in Two-Component Gene Circuit

C.-M. Ghim* and E. Almaas†
Microbial Systems Biology Group, Biosciences and Biotechnology Division,
Lawrence Livermore National Laboratory, 7000 East Avenue Livermore, CA 94550

Cell has the potential to remember the environmental conditions for many (10^7) generations but stochastic fluctuations set a fundamental limit on the stability of this memory. Here we explicitly take the binding-unbinding of macromolecules into account to propose a novel rationale for the protein-protein interaction in cell physiology. Based on the first-exit time and the corresponding deterministic characterization of various genetic circuits, we show that the reversible binding dynamics may stabilize non-genetically inherited cell states, providing a practical strategy for designing robust epigenetic memory.

PACS numbers: 87.18.Cf, 82.39.-k, 05.40.+j

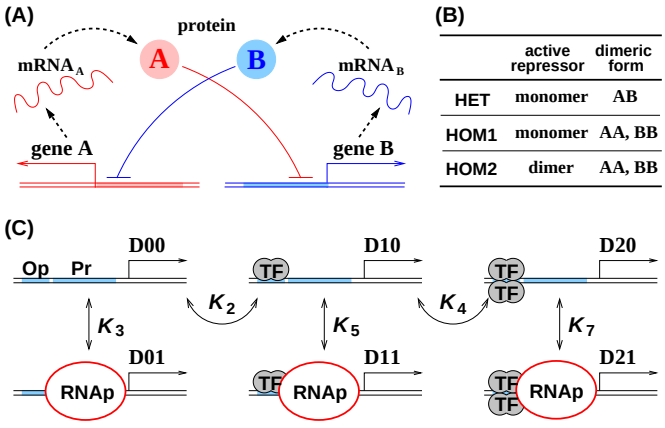
Introduction. Living cells are intrinsically noisy system. Only a few molecules synthesized in random burst can be sufficient to erratically turn on or off the gene expression, a problem that is only now being appreciated in electronic circuits as the size of the active elements shrinks to near-molecular dimensions. The transition between stable phenotypes can be driven not only by environmental cues but also by fluctuations due to random arrival of chemical reactions and low copy number of regulatory proteins [1]. Naturally, cells have evolved elaborate molecular machinery to get over this fundamental limit to the biochemical information processing or even to exploit it as a survival strategy. Recent experiments on a clonal population of microbes [2–4] have shown that the random switching does create phenotypic diversity, potentially giving a population better chance of survival as exemplified in bacterial persistence under antibiotic treatment [5] or in fluctuating environments in general [6, 7]. Noise-induced switching, however, more often signals a defect in cellular information processing. Untimely exit from latency means to viruses increased chance of being targeted by the host immune system. Transcriptional activation of sugar uptake genes in the wrong media may well be a severe waste of resources.

In fact, biochemical switches, such as *cI-cro* circuit of phase λ inside a bacterial cell [8], have been known for its exceptional stability. Unless induced by external agent (e.g. UV light), an individual phage in *E. coli* cell sustains its lysogenic reproduction mode throughout 10^7 generations [9, 10], which is lower than the genomic mutation rate of the host organism. Then the problem of random switching becomes a matter of balance: switching should be tunable balancing the two counteracting needs. Here we investigate tunable synthetic gene circuits, the dynamical capability of which is controlled by chemical specification of the constituent macromolecules. To quantitatively address the issue of stability against random switching and its dependence on the circuit topology and kinetic details as well, we employ the chemical master equation (CME) and Gillespie's numerical method [11] for the genetic toggle switch, where two genes transcriptionally repress each other's expression via their own protein products [12]. In addition to the trascription intitation control, we will look at the role of post-translational protein-protein binding dynamics as a handle for stabilizing gene expression states, which will find broad relevance to the efforts in synthetic biology.

Dynamics of genetic toggle switch. The stochastic model for a pair of mutually repressing genes ("A" and "B") comprises binding-undinding reactions among DNA, proteins, and RNA polymerases (RNAp). In physiological time scale for which functional macromolecules are synthesized or degraded, binding-unbinding reactions can be treated in an adiabatic manner. Then each reversible binding reaction is charaterized by a single thermodynamic constant, $K_i \equiv q_i/k_i$, where k_i and q_i is association and dissociation rate, respectively (Table 1). If we are able to neglect the stochastic fluctuations, the full dynamics of toggle switch can be reduced to the following dynamical system.

$$\begin{cases} \dot{x} = f_1(Y) - x - g_1(x,y) \\ \dot{y} = f_2(X) - y - g_2(y,x) \end{cases} \tag{1}$$

where x and y are now the concentration of protein species A and B scaled by the dissociation constant of protein-operator complex, K_2. Time is taken in units of the protein lifetime, which sets the timescale of cellular "memory" retention, and



	active repressor	dimeric form
HET	monomer	AB
HOM1	monomer	AA, BB
HOM2	dimer	AA, BB

FIG. 1: (Color online) (A) Circuitry of the genetic toggle switch. Broken lines are transcription and translation processes and the lines with blunt end denote the transcriptional repression. (B) Model genetic circuits under consideration. (C) Microscopic binding states of DNA. Op, operator; Pr, promoter; TF, transcription factor; RNAp, RNA polymerase. $\{K_i\}$ is the dissociation constant of each reaction.

Reaction	Rate	Reaction	Rate
P+P′ $\rightleftharpoons$ PP′	(k_1, q_1)	$\varnothing \rightarrow$ mRNA	α_m
D00+P $\rightleftharpoons$ D10	(k_2, q_2)	$\varnothing \rightarrow$ P	α_p
D00+RNAp $\rightleftharpoons$ D00	(k_3, q_3)	mRNA $\rightarrow \varnothing$	γ_m
D10+P $\rightleftharpoons$ D20	(k_4, q_4)	P $\rightarrow \varnothing$	γ_p
D10+RNAp $\rightleftharpoons$ D11	(k_5, q_5)		
D20+RNAp $\rightleftharpoons$ D21	(k_5, q_5)		

TABLE I: List of reactions. P denotes protein and P′ is either the same species or not. Other symbols are explained in Fig. 1.

is set the same for the two species. Upper case X and Y denotes the concentration of the active form of the repressor proteins derived from x and y, which can be either monomer or homo-/hetero-dimer.

The regulated synthesis rate f_i, often expressed as an *ad hoc* sigmoid with arbitrary Hill coefficients, is a monotone nonincreasing function. In the lumped circuit scheme, the gene expression rate is formulated as the product of promoter occupation probability by RNAp and the speed of the ensuing (mRNA or peptide) elogation. Likewise, since the transcriptional initiation control is mediated by binding the repressor proteins to the specific *cis*-regulatory DNA sequence, the problem of determining f_i is reduced to find the RNAp-promoter occupancy either in the absence of transcription factor (TF) or in its presence. We allow for the case, without losing genrality, where RNAp may bind to bare DNA or infrequently sit on the promoter site despite the steric hindrance by repressors, as depicted by the finite dissociation constants K_5 and K_7 in Fig. 1C. Then it proves that

$$f(X) = \lambda\left(1 + \frac{\nu}{1+\mu X + \mu r X^2}\right) , \qquad (2)$$

where $\lambda = f(\infty)$ is the dimensionless synthesis rate of proteins in full repression and $\nu = [f(0) - f(\infty)]/f(\infty)$ is the reduced fold change between ON and OFF states. Following mass-action kinetics, all these dimensionless compound parameters can be tracked into equilibrium thermodynamic constants:

$$\lambda = \frac{\beta}{1+q_3/s} ; \quad \mu = \frac{s+q_3}{1+q_3} ; \quad \nu = \frac{q_3(1-s)}{s(1+q_3)} ; \quad r = \frac{K_2}{K_4} ,$$

where $r > 1$ is the cooperativity factor in protein-DNA binding, $s \equiv K_3/K_5 \in (0,1)$ is the promoter "leakage", and $q_3 \equiv K_3/[\mathrm{RNAp}]$ is the RNAp-promoters dissociation constant scaled by the concentration of RNAp. A key parameter within experimental manipulation is the average protein copy number in the absence of regulation, $\beta \equiv \alpha_m\alpha_p/K_2\gamma_m\gamma_p$, where α_m (α_p) and γ_m (γ_p) respectively denotes the synthesis and degradation rate of mRNA (protein). The decay term other than the passive linear degradation due to cell growth reflects the effect of reversible binding-unbinding dynamics. Heterodimerization between repressor proteins dictates $g_i = xy/\kappa$ whereas the homodomer circuits have $g_i(x,y) = x^2/2\kappa$, where $\kappa \equiv \sigma K_1/K_2$ and σ is the ratio of dimer lifetime to monomer lifetime. K_1, the dissociation constant of protein dimer, constitutes the key

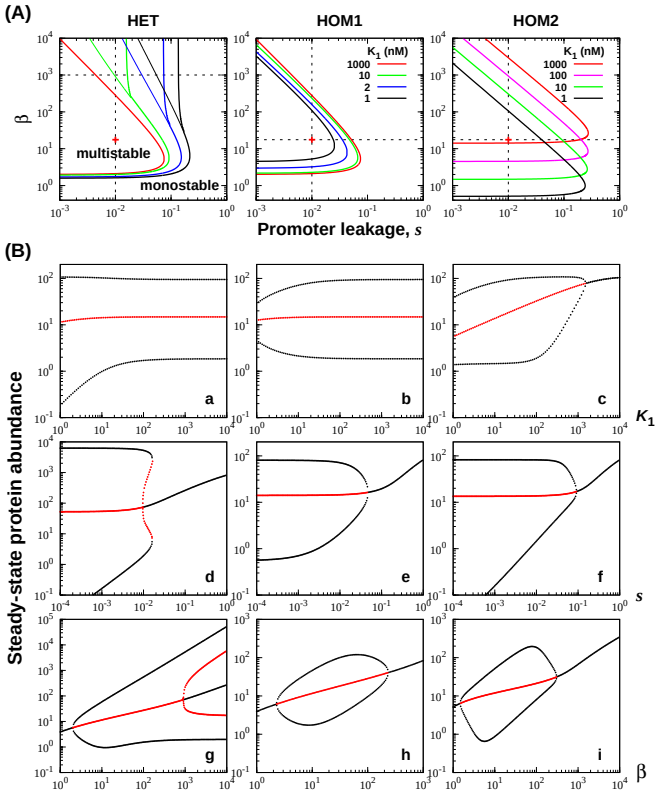


FIG. 2: (Color online) (A) Phase diagram of genetic toggle switches. The promoter leakage s and the mean protein copy number β in the absence of regulation are taken as two independent variables for a given series of K_1. Red cross symbol indicates the parameter set for stochastic simulations, and the steady-state molecular abundance (B) has been obtaind along the broken lines. Notice that for strong protein-protein binding affinity and high synthesis-loose repression limit, HET may develop higher multistability region (wedge-shaped area bounded by thick and thin lines), where three or more stable states are possible. (B) Steady-state protein abundance as a function of K_1 (top), s (middle), and β (bottom). From the fixed point in parameter space, $(K_1/\mathrm{nM}, s, \beta)$=(10, 0.01, 17.5 or 10^3 (HET)), each parameter is allowed to vary with the other two fixed. The trajectories in black (red) denote (un)stable steady states.

parameters along with s and β that can be manipulated in experiments.

If we consider the symmetric toggle switch for analytical simplicity, that is, the two species of proteins are subject to the same rate processes, the functional form of f_i and g_i is independent of i. Depending on the variable pathways of post-translational binding dynamics, we explore three different realization of the toggle switch: (i) HET model allows for the formation of inactive heterodimer. (ii) HOM1 and (iii) HOM2 model includes the homodimerization of each protein species, while the former (latter) has monomer (dimer) as the active form of transcription factor. As $K_1 \to \infty$, results from both (i) and (ii) converge to the the simplest monomer-only model, as is shown by the almost exact overlap of $K_1 = 1000$ nM phase boundaries in HET and HOM1 models (Fig. 2A). For the explicit form of the functions f in Eq. (1), the systems

with monomeric repressor (HET and HOM1) have $X = x$ and $Y = y$, while $X = \theta x^2$ and $Y = \theta y^2$ with $\theta \equiv K_2/K_1$ for dimeric repressor (HOM2).

Multistability controlled by s, β, and K_1. Depending on the kinetic rates involved in molecular binding/unbinding dynamics, Eq. (1) may develop bistability which provides a basis for biological switching devices. Null cline analysis gives the bistability region as shown in Fig. 2, where the promoter leakage and the chosen as independent parameters and all the others are fixed near the physiologically relevant values. A general tendency throughout the different model systems is that the decreased promoter leakage (small s) enhances the bistability. Also for the bistability, given fixed decay rates of mRNA and porteins, intermediate transcription-translation efficiency (β) is required. Both the excessive and depleted expression rates will bring the system to the globally unique steady state. Increased binding affinity in protein-protein interaction (smaller K_1) enhances the bistability region in HET circuit while suppressing it in HOM1. On the other hand, the bistable region of HOM2 circuit is parallel-shifted towards higher values of β, showing nonmonotonic behavior in the occurrence of random switching as K_1 varies (Fig. S1 in *Supplementary Information*). In both the weak and strong binding limit, HOM2 circuit displays frequent switching between steady states.

A penetrating explanation for all the observed behavior is the proximity effect near the phase boundary. As K_1 decreases, multistable region gets thicker (thinner) in HET (HOM1), leaving the system (red cross in Fig. 2A) more and more central (peripheral) within the multistable region. Because of the parallel shift in phase boundary, HOM2 circuit is placed in the middle of the bistable region only in the intermediate values of K_1. Around the physiological rates for the phage λ [13–20], all types of the toggle switch develop multistability in a wide range of parameter variations. The transitions from monostability to bistability or vice versa belong to the (supercritical) pitchfork bifurcation as shown in Fig. 2B, and so the dynamical system near those critical points shows sensitivity to small random perturbations due to the decreased separation of the two stable states in state space. For this bifurcations, the separation of the two attractors increases right after their appearance and then decreases back to merge with the the other attractor and a saddle point. On the other hand, the positive eigenvalue of the Jacobian at the saddle point (data not shown), which measures the "barrier height" between the two attractors, shows initial increase followed by the decrease to zero at the upper critical point.

Interestingly, HET circuit can be tuned to possess three or more stable states at the intermediate promoter leakage combined with the large ratio of synthesis to degradation rates of proteins. Starting from a bistable parameter region of small s and large β, as the transcriptional repression becomes looser (increasing s), a subcritical pitchfork bifurcation occurs to generate the third attractor and two extra saddle points (Fig. 2Bd). The same is true to the β scan (Fig. 2Bg). As s (β) keeps growing in Fig. 2Bd (Bg), HET turns to lose stable steady states to eventually restore monostability. In this process, the two attractors merge with the saddle points and annihilate each other (saddle-node bifurcation) to leave only a single attractor (not shown in Fig. 2Bg). Here unphysiologically high abundance of the dominant protein even in the moderate values of β is attributed to the depletion of minor species by binding with the dominant species.

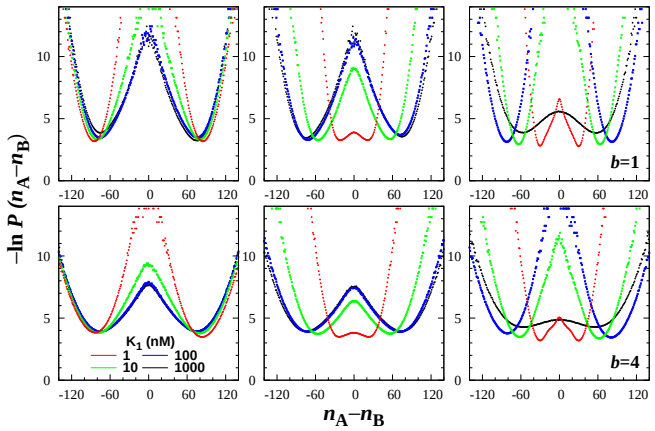


FIG. 3: (Color online) Effective potential reconstructed from the probability of the system to collapse into the line of constant $n_A - n_B$. For each circuit and each value of K_1, we ran two independent 10^6-generation-long simulations starting from symmetrically inverted initial conditions. Remnant asymmetry is the finite sampling effect. Two rows correspond to the burstiness of 1 and 4, respectively.

Stochastic approaches. To evaluate the stability of distinct toggle switches, the "barrier height" as welll as the separation of attractors should be considered. In the absence of the global "energy landscape", since the drift field of Eq. (1) is nonconservative, we need to resort to CME that empirically reveals the structure of the landscape. We generate the stochastic time series using Gillespie algorithm [11]. The bifurcation-based proximity effect persists in stochastic time courses (data not shown). Random switching in HET (HOM1) circuit occurs less (more) often for the stronger binding affinity between proteins, whereas HOM2 requires an intermediate range of binding affinity for the reduced switching rate. Dependence of the switching dynamics on the tunable characteristics of the toggle can also be captured by one-dimensional effective potential constructed by counting the occurrence of the difference $n_A - n_B$ (Fig. 3). Double-well landscape of $-\ln P(n_A - n_B)$ reconfirms the bifurcation analysis, where increased binding affinity between proteins leads to higher (lower) barrier height and larger (smaller) separation of the two attractor for HET (HOM1) while the trend is mixed in HOM2. It has been known [21] that the average number of translated proteins from a single mRNA template during mRNA lifetime ($\alpha_p/\gamma_m \equiv b$) plays an important role in controlling the intrinsic noise. In deterministic dynamics, however, this rate is lumped with the number of transcription per cell cycle (α_m/γ_p) to give an effective system parameter β. To address the issue of the translational burst, we separate stoichiometry from propensity

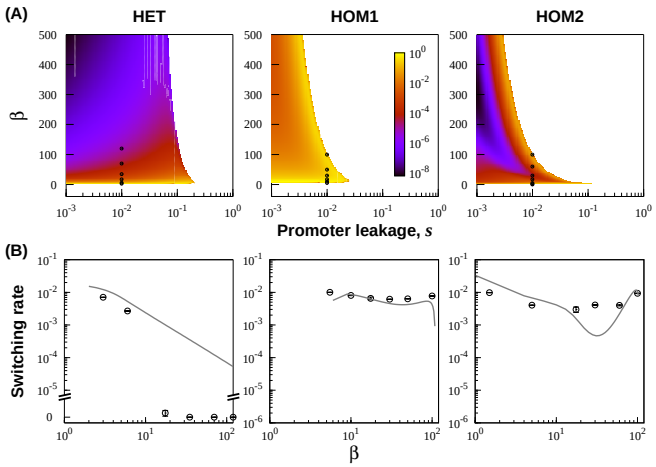


FIG. 4: (Color online) (A) Heat map of mean first-passage time calculated by the approximation scheme explained. (B) Empirical switching frequencies obtained from stochastic time series (black dots) are compared with the approximate numerical results. $K_1 = 1$ nM, $b = 4$, $s = 0.01$. All the rates are displayed in units of hour^{-1}.

in stochastic approach, and factorize λ into the "burstiness" b and the remaining factors.

Calculation of switching rate. In linear noise approximation [22] the fluctuation of protein copy numbers around the macroscopic concentration $\boldsymbol{r} = (x,y)$ is given by $\boldsymbol{n} - V_c\boldsymbol{r} \equiv \sqrt{V_c}\boldsymbol{\xi}$ (V_c is the cell volume). Using Kramers-Moyal expansion of the discrete step operator, stochastic dynamics for $\boldsymbol{n}$ can be approximated by Fokker-Planck equation for the continuous random variable $\boldsymbol{\xi}$.

$$\dot{P}(\boldsymbol{\xi},t) = -J_{ij}\frac{\partial}{\partial\xi_i}\left[\xi_j P(\boldsymbol{\xi},t)\right] + \frac{1}{2}D_{ij}\frac{\partial^2}{\partial\xi_i\partial\xi_j}P(\boldsymbol{\xi},t)\ , \qquad (3)$$

where $i = 1,2$ and the summation convention is used for repeated indices. The drift field $J_{ij}\xi_j$ is the linearization of the right-hand side of Eq. (1), and the diffusion matrix explicitly reveals the dependence on b, i.e.

$$J_{ij} = \begin{pmatrix} -1-g_x(x,y) & f_y(Y) \\ f_x(X) & -1-g_y(y,x) \end{pmatrix}; \qquad (4a)$$

$$D_{ij} = \begin{pmatrix} bf(Y)+x+g(x,y) & 0 \\ 0 & bf(X)+y+g(y,x) \end{pmatrix}, \qquad (4b)$$

where (x,y) and (X,Y) are taken from the deterministic trajectory, and subscripts indicate the partial derivative with respect to the index designated. With this preparation, we quantitate the stability of the different toggle switches in relation to the first-passage time in large deviation theory [23]. Thus obtained switching rates are visualized in Fig. 4. As the deterministic analysis has previously shown, toggle switch is more stabilized against the random fluctuation when the system parameters are located more inside the multistability region (Fig. 4).

Conclusions. A key challenge in the emerging area of synthetic biology is to identify general, scalable strategies that enable construction of increasingly complex gene circuits with reliable performance. We comparatively evaluated different realization of genetic toggle switch that potentially control the gene expression in further integrated genetic regulatory circuits. We observed the nontrivial dependence of the intrinsic noise on protein-DNA interactions and post-translational protein-protein binding dynamics as well. In particular, HET and HOM2 outperform HOM1 in the stability of the memory, providing a candidate design for robust switching or memory devices. A more important aspect is the tunability of the switching dynamics that may well have driven living organisms into the selective pressure and now provides engineering targets for technological purpose as well. For instance, to reduce the promoter leakage s, the operator sequence can be engineered to accommodate extra protein-binding sites. In order to increase (decrese) β, one can increase (decrease) the synthesis rates of mRNA/protein and/or decrease (increase) the degradation rate thereof. A practical way to achieve this goal is to fuse the ssRA tag into the 3' end of each gene [24] so that the protein degradation is exposed to active proteolysis. Fianlly, K_1 can be modulated by modifying the binding domain of the repressor proteins. A novel avenue for this direction includes the integration of the evolutionary instabilities into the construction principles of engineered biological systems [25].

This work was performed under the auspices of the U. S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This project (06-ERD-061) was funded by the Laboratory Directed Research and Development Program at LLNL.

* Electronic address: cmghim@llnl.gov
† Electronic address: almaas@llnl.gov

[1] P. Guptasarma, Bioessays **17**, 987 (1995).
[2] J. Newman, S. Ghaemmaghami, J. Ihmels, D. Breslow, M. Noble, J. DeRisi, and J. Weissman, Nature **441**, 840 (2006).
[3] N. Rosenfeld, J. Young, U. Alon, P. Swain, and M. Elowitz, Science **307**, 1962 (2005).
[4] J. Pedraza and A. van Oudenaarden, Science **307**, 1965 (2005).
[5] N. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler, Science **305**, 1622 (2004).
[6] E. Kussel and S. Leibler, Science **309**, 2075 (2005).
[7] M. Acar, J. Mettetal, and A. van Oudenaarden, Nature Genetics **40**, 471 (2008).
[8] M. Ptashne, *A Genetic Switch: Gene Control and Phage* λ (Cambridge, MA, 1986).
[9] D. Rozanov, R. D'Ari, and S. Sineoky, J Bacteriology **180**, 6306 (1998).
[10] J. Little, D. Shepley, and D. Wert, EMBO Journal **18**, 4299 (1999).
[11] D. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[12] T. Gardner, C. Cantor, and J. Collins, Nature **403**, 339 (2000).
[13] R. Sauer, Ph.D. thesis, to: Dept. of Biochemistry and Molecular Biology, Harvard University (1979).
[14] A. Arkin, J. Ross, and H. McAdams, Genetics **149**, 1633 (1998).
[15] G. Ackers, A. Johnson, and M. Shea, Proc Natl Acad Sci USA

79, 1129 (1982).
[16] D. Hawley and W. McClure, J Mol Biol **157**, 493 (1982).
[17] D. Hawley, A. Johnson, and W. McClure, J Biol Chem **260**, 8618 (1985).
[18] P. L. deHaseth, M. L. Zupancic, and M. T. Record, J Bacteriol **180**, 3019 (1998).
[19] A. Ujvari and C. Martin, Biochem **35**, 14574 (1996).
[20] M. Shea and G. Ackers, J Mol Biol **181**, 211 (1985).
[21] M. Thattai and A. van Oudenaarden, Proc. Natl. Acad. Sci. U.S.A. **98**, 8614 (2001).
[22] N. van Kampen, *Stochastic processes in physics and chemistry* (Elsevier North-Holland, New York, 1981).
[23] M. Freidlin and A. Wentzell, *Random perturbations of dynamical systems, vol. 260 of Grundlehren der Mathematischen Wissenschaften* (Springer, New York, NY,, 1998).
[24] C. Grilly, J. Stricker, W. Pang, M. Bennett, and J. Hasty, Mol. Sys. Biol. **3** (2007).
[25] E. Haseltine and F. Arnold, Annu. Rev. Biophys. Biomol. Struct. **36**, 1 (2007).